

On the Role of Public and Private Assessments in Security Information Sharing Agreements

Parinaz Naghizadeh and Mingyan Liu

Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor
{naghizad, mingyan}@umich.edu

In recent years, sharing of security information among organizations, particularly information on both successful and failed security breaches, has been proposed as a method for improving the state of cybersecurity. However, there is a conflict between individual and social goals in these agreements: despite the benefits of making such information available, the associated disclosure costs (e.g., drop in market value and loss of reputation) act as a disincentive for firms' full disclosure. In this work, we take a game theoretic approach to understanding firms' incentives for disclosing their security information given such costs. We propose a repeated game formulation of these interactions, allowing for the design of inter-temporal incentives (i.e., conditioning future cooperation on the history of past interactions). Specifically, we show that a rating/assessment system can play a key role in enabling the design of appropriate incentives for supporting cooperation among firms. We further show that in the absence of a monitor, similar incentives can be designed if participating firms are provided with a communication platform, through which they can share their beliefs about others' adherence to the agreement.

Key words: Information sharing agreements, breach disclosure, repeated games, inter-temporal incentives

1. Introduction

Improving the ability of analyzing cyber-incidents, and ensuring that the results are shared among organizations and authorities in a timely manner, has received increased attention in the recent years by governments and policy makers, as it can lead to a better protection of the national infrastructure against potential cyber-attacks, allow organizations to invest in the most effective preventive and protective measures, and protect consumer rights.

In the US, improving information sharing is listed as one of President Obama's administration's priorities on cybersecurity, and is evidenced by its inclusion as one of the key focus areas in the 2013 Executive Order 13636 on "Improving Critical Infrastructure Cybersecurity" (Obama (2013)), and as initiative #5 in the Comprehensive National Cybersecurity Initiative (CNCI) (The White House (2015b)). Most recently, during the first White House Summit on cybersecurity and consumer protection, President Obama signed Executive Order 13691 on "Promoting Private Sector Cybersecurity Information Sharing", encouraging companies to share cybersecurity information with one another and the federal government (Obama (2015)). Following the executive order, the Department of Homeland Security (DHS) has started efforts to encourage the development of Information

Sharing and Analysis Organizations (ISAOs) (DHS (2015b)), as well as the Cyber Information Sharing and Collaboration Program (CISCP) in order to encourage Cooperative Research and Development Agreements. As of July 2015, 125 such agreements have been placed, with an additional 156 being negotiated (The White House (2015a)).

In general, depending on the breach notification law or the information sharing agreement, a firm may be required to either publicly announce an incident, to report it to other firms participating in the agreement or within its industry sector, to notify affected individuals, and/or to notify the appropriate authorities. Currently, most of the existing laws in the US and the European Union require organizations to only report to an authority, with a few other also mandating notification of the affected individuals; e.g., HIPAA for the health sector in the US (see Laube and Böhme (2015) for a summary of prominent US and EU laws). However, motivated by the aforementioned trend in the newest initiatives in the US (in particular, EO 13691), in this paper, we are primarily interested in information sharing agreements *among firms*, both with and without facilitation by an authority. Examples of existing agreements/organizations of this type include Information Sharing and Analysis Centers (ISACs), Information Sharing and Analysis Organizations (ISAOs), the United States Computer Emergency Readiness Team (US-CERT), and InfraGard. Currently, joining and reporting in all such information sharing organizations is voluntary.

1.1. Problem Motivation

Several studies have analyzed the positive effects of information sharing laws. Romanosky et al. (2011) show that the introduction of breach disclosure laws has resulted in a reduction in identity theft incidents. Gordon et al. (2015) argue that shared information can reduce the uncertainty in adopting a cybersecurity investment, thus leading firms to take a proactive rather than reactive approach to security, and consequently increasing the expected amount of investments in cybersecurity. Finally, Gordon et al. (2006) show that the Sarbanes-Oxley Act of 2002 (despite only indirectly encouraging higher focus on reporting of security-related information) has had a positive effect on disclosure of information security by organizations. Nevertheless, there exist anecdotal and empirical evidence that security breaches remain under-reported, see e.g., Claburn (2008), Threat Track (2013).

These observed disincentives by companies for sharing security information can be primarily explained by analyzing the associated economic impacts. Campbell et al. (2003), Cavusoglu et al. (2004) conduct event-study analyses of market reaction to breach disclosures, both demonstrating a drop in market values following the announcement of a security breach. In addition to an initial drop in stock prices, an exposed breach or security flaw can result in loss of consumer/partner confidence in a company, leading to a further decrease of revenues in the future (Gal-Or and Ghose

(2005)). Finally, documenting and announcing security breaches impose a bureaucratic burden on the company; e.g, when an agreement requires the reports to comply with a certain incident reporting terminology; examples of such frameworks include the recently proposed categorization by DHS (DHS (2015a)), and the Vocabulary for Event Recording and Incident Sharing (VERIS) proposed by the Verizon RISK team (Verizon (2015)).

Given these potential disclosure costs, and the evidence of under-reporting of security information, it is clear that we need a better understanding of firms' incentives for participating in information sharing organizations, as well as the economic incentives that could lead to voluntary cooperation by firms in these agreements.

1.2. Related Work

A number of research papers have analyzed the welfare implications of information sharing agreements, as well as firms' incentives for adhering to these agreements.

The work by Ogut et al. (2005) and Laube and Böhme (2015) consider the effects of security breach reporting between firms and an authority. Ogut et al. (2005) show that if the availability of shared information¹ can reduce either attack probabilities or firms' interdependency, it will benefit social welfare by inducing firms to improve investments in self-protection and cyber-insurance. On the other hand, Laube and Böhme (2015) study the effectiveness of mandatory breach reporting, and shows that enforcing breach disclosure to an authority (through the introduction of audits and sanctions) is effective in increasing social welfare only under certain conditions, including high interdependence among firms and low disclosure costs.

Gordon et al. (2003) and Gal-Or and Ghose (2005) propose game-theoretic models of information sharing among firms. Gordon et al. (2003) show that, if security information from a partner firm is a *substitute* to a firm's own security expenditures, then (mandatory) information sharing laws reduce expenditure in security measures, but can nevertheless increase social welfare. However, firms will not voluntarily comply with sharing agreements, requiring additional economic incentives to be in place (e.g., a charge on a member of the ISAC for losses on the other member). Gal-Or and Ghose (2005) on the other hand allow information sharing to be a *complement* to the firm's own security expenditures, as it may increase consumer confidence in a firm that is believed to take steps towards securing her system. Using this model, the authors show that when the positive demand effects of information sharing are high enough, added expenditure and/or sharing by one firm can incentivize the other firm to also increase her expenditure and/or sharing levels.

¹ Firms' incentives for information disclosure or the mechanisms for ensuring breach disclosure have not been modeled in Ogut et al. (2005).

In this work, similar to Gordon et al. (2003), Ogut et al. (2005), Laube and Böhme (2015), we assume disclosure costs are higher than potential demand-side benefits, therefore similarly predicting a lack of voluntary information sharing at the state of equilibrium. Our proposed approach of considering the effects of repeated interactions as an incentive solution is however different from those proposed in aforementioned literature, as they consider one-shot information sharing games.

1.3. Inter-temporal Incentives in Information Sharing Agreements

In this paper, we also take a game-theoretic approach to understand firms' behavior and (dis)incentives in security information sharing agreements. This approach is motivated by the fact that despite the aforementioned disclosure costs (which deter firms from joining such agreements and sharing their security information), full disclosure has benefits for participating firms, as each firm can prevent similar attacks and invest in the best security measures by leveraging other firms' experience. Consequently, an outcome in which firms disclose their information would be preferred by all participants. There is therefore a conflict between individual interest and societal goals. To capture this conflict, we model security information sharing agreements as an N-person prisoner's dilemma (NPD) game. In an NPD, there will be no information sharing at the state of equilibrium, as also predicted by similar game-theoretic models which consider one-shot information sharing games (see Section 1.2). Existing research has further proposed audits and sanctions (e.g. by an authority or the government) or introducing additional economic incentives (e.g. taxes and rewards for members of ISACs) as remedies for encouraging information disclosure.

In this paper, we take a different approach and account for the repeated nature of these agreements to propose the design of inter-temporal incentives that lead sufficiently patient firms to cooperate on information sharing. It is well-known in the economic literature that repetitions of an otherwise non-cooperative and inefficient game can lead economically rational agents to coordinate on efficient equilibria, see Mailath and Samuelson (2006). The possibility of achieving efficient outcomes however depends on whether the monitoring of other participants' actions is perfect or imperfect, and private or public. In particular, for information sharing games, each firm or an outside monitor can (at best) only imperfectly assess the honesty and comprehensiveness of the shared information. Accordingly, we consider two possible monitoring structures for these games.

First, we analyze the role of a rating/assessment system in providing an imperfect public signal about the quality of firms' reports in the agreement. We show that for the proposed NPDs equipped with a simple monitoring structure, the folk theorem of Fudenberg et al. (1994) holds in the repeated game, therefore making it possible to design appropriate inter-temporal incentives to support cooperation. We illustrate the construction of these incentives through an example, and discuss the effect of the monitoring accuracy on this construction. We then consider the design

of such incentives in the absence of a public monitoring system. Specifically, we assume that the firms have access to a communication platform, through which they are allowed to report their private beliefs on whether other firms are adhering to the agreement. We show that given a simple imperfect private monitoring structure by each firm, the folk theorem of Kandori and Matsushima (1998) will be applicable to our proposed NPDs, again enabling the design of appropriate incentives for information sharing.

Contributions. The contributions of the current work are therefore the following:

- We propose the design of inter-temporal incentives for supporting cooperative behavior in security information sharing agreements. To this end, we model firms' interactions as an N-person prisoner's dilemma game and equip it with a simple monitoring structure.
- We illustrate the role of a public rating/assessment system in providing imperfect public monitoring, leading to coordination on cooperation in information sharing agreements.
- We further establish the possibility of sustaining cooperative behavior in the absence of a public monitor, by introducing a platform for communication among firms.

Preliminary versions of this work appeared in Naghizadeh and Liu (2016a,b). We first proposed the idea of using inter-temporal incentives in information sharing agreements in Naghizadeh and Liu (2016a). We analyzed the possibility of using public monitoring in a two-person prisoner's dilemma game in Naghizadeh and Liu (2016b). In the current work, we generalize the model to N-person prisoner's dilemma games. We extend our analysis of public monitoring to the general model, and further analyze the possibility of using private monitoring through an appropriate communication platform.

Paper Organization. We present the model and proposed monitoring structure for information sharing games in Section 2. We discuss the role of a monitoring system in the design of inter-temporal incentives in Section 3, followed by the analysis of providing such incentives based on private observations and communication among firms in Section 4. We conclude in Section 5.

2. Information Sharing Game Model

2.1. The Stage Game

Consider N (symmetric) firms participating in an information sharing agreement (e.g. firms within an ISAC). Each firm can choose a level of expenditure in security measures to protect her infrastructure against cyber incidents. Examples include implementing an intrusion detection system, introducing employee education initiatives, and installing and maintaining up-to-date security software. We assume these measures are implemented independently of the outcome of the sharing agreement, and focus solely on firms' information sharing decisions.²

² This assumption is adapted for two reasons. First, this allows us to focus only on firms' incentives for information sharing. More importantly, here the information shared by firm i is assumed to be a *substitute* to firm j 's investment;

The information sharing agreement requires each firm i to share her security information with other participating firms. This disclosure can include information on both successful and failed attacks, as well as effective breach prevention methods and the firm's adopted security practices. A firm i should therefore decide whether to fully and honestly disclose such information. We denote the decision of firm i by $r_i \in \{0, 1\}$, with $r_i = 0$ denoting (partially) concealing and $r_i = 1$ denoting (fully) disclosing.³ Denote the number of firms adopting a full disclosure decision of by x ; i.e., $x := |\{i \mid r_i = 1\}|$.

A decision of $r_i = 1$ is beneficial for the following reasons. On one hand, the disclosed information can allow other firms $j \neq i$ to leverage the acquired information to protect themselves against ongoing attacks and to adopt better security practices. Aside from this security-related implications, information disclosure $r_i = 1$ may further provide a competitive advantage to firms $j \neq i$, allowing a firm j to increase her share of the market by strategically leveraging the attained information to attract a competitor i 's costumers. Finally, sharing of security information may be beneficial to firm i herself as well (especially when many other firms are disclosing as well), as it may garner trust from potential partners and costumers. We denote all such applicable *information gains* to a firm, as a function of firm i 's decision and the number of *other* firms making a full disclosure decision, by $G(r, z) : \{0, 1\} \times \{0, 1, \dots, N - 1\} \rightarrow \mathbb{R}_{\geq 0}$, with $G(0, 0) = 0$. We assume that given r , $G(r, \cdot)$ is increasing in z , the number of other firms sharing their information.

Despite the aforementioned benefits of adopting $r_i = 1$, firm i has a disincentive for full disclosure due to the associated costs. These costs includes the man-hours spent in documenting and reporting security information, as well as potential losses in reputation, business opportunities with potential collaborators, stock market prices, and the like, following the disclosure of a breach or existing security flaws. In addition, it may be in i 's interest to conceal methods for preventing ongoing threats, predicting that an attack on the competitor j will result in j 's costumers switching to i 's products/services, increasing firm i 's profits. Consequently, such potential market loss or competitor's gain in sales can further deter firms from adhering to information sharing agreements. We denote all these associated *disclosure costs* by $L(r, z) : \{0, 1\} \times \{0, \dots, N - 1\} \rightarrow \mathbb{R}^+$, with $L(0, 0) = 0$, where the cost can potentially depend on how many other firms, z , are disclosing their security information.

i.e., firm j can decrease her security expenditure when she receives information from firm i . This possible reduction in the positive externality from j 's investments may therefore result in further disincentives for firm i for sharing her security information. We therefore remove these effects by decoupling the decisions and assuming fixed security expenditures. Analyzing the interplay of investment and sharing decisions remains a direction of future work.

³ The results and intuition obtained in the following sections continue to hold when firms can choose one of finitely many disclosure levels, given an appropriate extension of utilities and the monitoring structure.

We can now define the utility of each user based on her disclosure decision. Given the number of firms that are sharing, x , and substituting $z = x - \mathbb{1}\{r_i = 1\}$ (where $\mathbb{1}(\cdot)$ denotes the indicator function), we define the following utilities for the cooperators ($r_i = 1$) and deviators ($r_i = 0$):

$$\begin{aligned} \text{Cooperator:} \quad C(x) &:= G(1, x-1) - L(1, x-1) , \\ \text{Deviator:} \quad D(x) &:= G(0, x) - L(0, x) . \end{aligned}$$

We impose the following two assumptions on the utility functions:

ASSUMPTION 1. *Non-cooperation dominates cooperation,*

$$(A1) \quad D(x-1) > C(x), \quad \forall 1 \leq x \leq N .$$

Assumption (A1) entails that the disclosure costs outweigh the gain from sharing for the firm, making $r_i = 0$ a dominant strategy. In other words, the marginal benefit from increased trust or approval due to disclosure is limited compared to the potential market and reputation loss due to disclosed security weaknesses. Therefore, the only Nash equilibrium of a one-shot information sharing game is for no firm to disclose her information. This observation is consistent with similar studies of one-shot information sharing games in Gordon et al. (2003), Laube and Böhme (2015), which also conclude that, in the absence of audit mechanisms or secondary incentives, firms will choose to share no information because of the associated disclosure costs.

ASSUMPTION 2. *Non-cooperation is inefficient,*

$$(A2) \quad C(N) > D(0) = 0 .$$

Assumption (A2) entails that the resulting non-disclosure equilibrium is suboptimal, particularly compared to the outcome in which all firms disclose. That is, full disclosure dominates the unique Nash equilibrium of the one-shot game. We may further be interested in imposing a more restrictive condition (although this is not necessary for our technical discussion).

$$(A2') \quad xC(x) + (N-x)D(x) > (x-1)C(x-1) + (N-x+1)D(x-1), \quad \forall 1 \leq x \leq N .$$

Under (A2') (which indeed implies (A2)), non-disclosure by any firm decreases social welfare, making the full disclosure equilibrium $x = N$ the socially desired outcome.

EXAMPLE 1. Consider the gain functions $G(1, z) = G(0, z) = zG$ and loss functions $L(0, z) = 0$ and $L(1, z) = L$. Here, each firm obtains a constant gain G from any other firm who is disclosing information, and incurs a constant loss L if she discloses herself, both regardless of the number of other firms making a disclosure decision. It is easy to verify that these functions satisfy (A1). Furthermore, if $G > \frac{L}{N-1}$, assumptions (A2) and (A2') hold as well. Note also that the 2-player prisoner's dilemma can be recovered as a special case when $N = 2$.

EXAMPLE 2. Alternatively, consider the gain functions $G(1, z) = G(0, z) = f(z)G$, where $f(\cdot) : \{0, \dots, N-1\} \rightarrow \mathbb{R}^+$ is an increasing and concave function, and loss functions $L(0, z) = 0$ and $L(1, z) = L$. The concavity of $f(\cdot)$ implies that as the number of cooperators increases, the marginal increase in information gain is decreasing due to potential overlap in the disclosed information. The utilities of cooperators and deviators will be given by:

$$C(x) = f(x-1)G - L, \text{ and, } D(x) = f(x)G .$$

Assumption (A1) follows. Assumptions (A2) will hold if and only if $G > \frac{L}{f(N-1)-f(0)}$. However, unlike the previous example, for (A2') to hold we need additional restrictions beyond that required for (A2). Specifically, the full disclosure equilibrium will be the optimal solution only if the constants G and L are such that:

$$G[(N-x)(f(x) - f(x-1)) + (x-1)(f(x-1) - f(x-2))] > L, \forall x .$$

The described N -player game with assumptions (A1) and (A2) (or (A2')) is known as the N -person Prisoner's Dilemma (NPD) game; see e.g., Bonacich et al. (1976), Goehring and Kahan (1976). These games are used to model social situations in which there is a conflict between individual and societal goals; e.g., individual decisions whether to belong to unions, political parties, or lobbies, and problems of pollution or overpopulation (Bonacich et al. (1976)). The imposed assumptions then model the intuition that in such situations, any individual has a disincentive for cooperation, (A1), despite the fact that an outcome in which all cooperate would have been preferred by each participant, (A2).

2.2. Repeated Interactions and the Monitoring Structure

Throughout the following sections, we are interested in the design of inter-temporal incentives that can incentivize firms to move away from the one-shot equilibrium of the information sharing game, and adopt disclosure decisions when they interact repeatedly. Such inter-temporal incentives should be based on the history of firms' past interactions. We therefore formalize firms' monitoring capabilities, and the ensuing beliefs, of whether other firms are adhering to the information sharing agreement.

First, note that such monitoring is inevitably *imperfect*; after all, the goal of an information sharing agreements is to encourage firms to reveal their non-verifiable and private breach and security information. Furthermore, the monitoring can be either carried out independently by the firms, or be based on the reports of a central monitoring system. We consider both possibilities.

2.2.1. Imperfect Private Monitoring. First, assume each firm conducts her own monitoring and forms a belief on other firms' disclosure decisions. Specifically, by monitoring firm j 's externally observed security posture, firm i forms a *belief* b_{ij} about j 's report. We let $b_{ij} = 1$ indicate a belief by firm i that firm j has been honest and is fully disclosing all information, and $b_{ij} = 0$ otherwise. In other words, $b_{ij} = 0$ indicates that firm i 's monitoring provides her with evidence that firm j has experienced an undisclosed breach, has an unreported security flaw, or has fabricated an incident. Formally, we assume the following distribution on firm i 's belief given firm j 's report:

$$\pi(b_{ij}|r_j) = \begin{cases} \epsilon, & \text{for } b_{ij} = 0, r_j = 1 \\ 1 - \epsilon, & \text{for } b_{ij} = 1, r_j = 1 \\ \alpha, & \text{for } b_{ij} = 0, r_j = 0 \\ 1 - \alpha, & \text{for } b_{ij} = 1, r_j = 0 \end{cases} \quad (1)$$

with $\epsilon \in (0, 1/2)$ and $\alpha \in (1/2, 1)$. First, note that ϵ is in general assumed to be small; therefore, if firm j fully discloses all information ($r_j = 1$), firm i 's belief will be almost consistent with the received information. Intuitively, this entails the assumption that with only a small probability ϵ , firm i will be observing flaws or breaches that have gone undetected by firm j herself, as internal monitoring is more accurate than externally available information. On the other hand, firm i has an accuracy α in detecting when firm j conceals security information ($r_j = 0$). Note that $(\epsilon = 0, \alpha = 1)$ is equivalent to the special case of perfect monitoring.

We assume the evidence available to firm i , and hence the resulting belief b_{ij} , is private to firm i , and independent of all other beliefs. Specifically, $b_{ij}, \forall i \neq j$ are i.i.d. samples of a Bernoulli random variable (with parameter α or ϵ depending on r_j).

2.2.2. Imperfect Public Monitoring. Alternatively, consider an independent entity (the government, a white hat, or a research group), referred to as *the monitor*, who assesses the comprehensiveness of firms' disclosure decisions, and publicly reveals the results. We assume the distribution of the beliefs $\{b_{01}, \dots, b_{0N}\}$ formed by the monitor is:

$$\hat{\pi}(\{b_{01}, \dots, b_{0N}\}|\{r_1, \dots, r_N\}) := \prod_{j=1}^N \pi(b_{0j}|r_j), \quad (2)$$

where the distributions $\pi(b_{0j}|r_j)$ follow (1), with ϵ and α interpreted similarly. Note that the monitoring technology of the monitor, i.e. (α, ϵ) , may in general be more accurate than that available to the firms.⁴

⁴It is worth mentioning that the binary beliefs are assumed for ease of exposition; the results of the subsequent sections continue to hold if the monitoring technology has finitely many outputs.

3. Imperfect Public Monitoring: The Role of Centralized Monitoring

The possibility of public monitoring (either perfect or imperfect) can enable the design of intertemporal incentives for cooperation in repeated interactions. With perfect public monitoring, deviations from the intended equilibrium path are perfectly observable by all participants, and can be accordingly punished. As a result, it is possible to design appropriate punishment phases (i.e., a finite or infinite set of stage games in which deviators receive a lower payoff) that keep sufficiently patient players from deviating to their myopic (stage game) best responses. This has led to folk theorems under perfect monitoring; see e.g., Fudenberg and Maskin (1986). With imperfect public monitoring on the other hand, deviations can not be detected with complete certainty. Nevertheless, the publicly observable signals can be distributed so that some are more indicative that a deviation has occurred. In that case, as players can all act based on their observations of the same signal to decide whether to start punishment or cooperation phases, despite the fact that punishment phases may still occur on the equilibrium path, it is possible for the players to cooperate to attain higher payoffs than those of the stage game.

In the remainder of this section, we first formalize the above intuition by presenting some preliminaries on infinitely repeated games with imperfect public monitoring, and in particular, the folk theorem of Fudenberg et al. (1994) for these games. In Section 3.2, we show that this folk theorem applies to NPD information sharing games with monitoring given by (2).

3.1. The Folk Theorem with Imperfect Public Monitoring

In this section, we present the folk theorem due to Fudenberg et al. (1994). Consider N rational players. At the stage game, each player i chooses an action $r_i \in R_i$. Let $\mathbf{r} \in R := \prod_{i=1}^N R_i$ denote a profile of actions. At the end of each stage, a public outcome $b \in B$ is observed by all players, where B is a finite set of possible signals. The realization of the public outcome b depends on the profile of actions \mathbf{r} . Formally, assume the probability of observing b following \mathbf{r} is given by $\pi(b|\mathbf{r})$. Let $u_i^*(r_i, b)$ be the utility of player i when she plays r_i and observes the signal b . Note that i 's utility depends on others' actions only through b , and thus the stage payoffs are not informative about others' actions. The ex-ante stage game payoff for user i when \mathbf{r} is played is given by:

$$u_i(\mathbf{r}) = \sum_{b \in B} u_i^*(r_i, b) \pi(b|\mathbf{r}). \quad (3)$$

Let \mathcal{F}^\dagger denote the set of convex combinations of players' payoffs for outcomes in R , i.e., the convex hull of $\{(u_1(\mathbf{r}), \dots, u_n(\mathbf{r})) | \mathbf{r} \in R\}$. We refer to \mathcal{F}^\dagger as the set of *feasible* payoffs. Of this set of payoffs, we are particularly interested in those that are *individually rational*: an individually rational payoff

profile \mathbf{v} is one that gives each player i at least her minmax payoff $\underline{v}_i := \min_{\boldsymbol{\rho}_{-i}} \max_{r_i} u_i(r_i, \boldsymbol{\rho}_{-i})$ (where $\boldsymbol{\rho}_{-i}$ denotes a mixed strategy profile by players other than i). Let $\boldsymbol{\rho}^i$, with

$$\begin{aligned}\boldsymbol{\rho}_{-i}^i &:= \arg \min_{\boldsymbol{\rho}_{-i}} \left(\max_{r_i} u_i(r_i, \boldsymbol{\rho}_{-i}) \right), \\ \rho_i^i &:= \max_{r_i} u_i(r_i, \boldsymbol{\rho}_{-i}^i),\end{aligned}$$

denote the minmax profile of player i , and $\mathcal{F}^* := \{\mathbf{v} \in \mathcal{F}^\dagger | v_i > \underline{v}_i, \forall i\}$ denote the set of feasible and strictly individually rational payoffs. The main purpose of a folk theorem is to specify which of the payoffs in \mathcal{F}^* (of which Pareto efficient payoffs are of particular interest) can be supported (as average payoffs) by some equilibrium of the repeated game.

Let us now discuss the repeated game. When the stage game is played repeatedly, at time t , each player has a private history containing her own past actions, $h_i^{t-1} := \{r_i^0, \dots, r_i^{t-1}\}$, as well as a public history of the public signals observed so far, $h^{t-1} := \{b^0, \dots, b^{t-1}\}$. Player i then uses a mapping σ_i^t from (h_i^{t-1}, h^{t-1}) to (a probability distribution over) R_i to decide her next play. We refer to $\sigma_i = \{\sigma_i^t\}_{t=0}^\infty$ as player i 's strategy. Each player discounts her future payoffs by a discount factor δ . Hence, if player i has a sequence of stage game payoffs $\{u_i^t\}_{t=0}^\infty$, her average payoff throughout the repeated game is given by $(1 - \delta) \sum_{t=0}^\infty \delta^t u_i^t$. Player is choosing her strategy σ_i to maximize this expression.

Among the set of all possible strategies σ_i , we will consider *public strategies*: these consist of decisions σ_i^t that depend only on the public history h^{t-1} , and not on player i 's private information h_i^{t-1} . Whenever other players are playing public strategies, then player i will also have a public strategy best-response. A *perfect public equilibrium (PPE)* is a profile of public strategies that, starting at any time t and given any public history h^{t-1} , form a Nash equilibrium of the game from that point on. PPEs facilitate the study of repeated games to a great extent, as they are “recursive”. This means that when a PPE is being played, the continuation game at each time point is strategically isomorphic to the original game, and therefore the same PPE is induced in the continuation game as well. Note that such recursive structure can not be recovered using private strategies, leading to the comparatively limited results in private monitoring games, as discussed in Section 4. Let $\mathcal{E}(\delta)$ be the set of all payoff profiles that can be attained using public strategies as PPE average payoffs when the discount factor is δ . We know that $\mathcal{E}(\delta) \subseteq \mathcal{F}^*$. The main question is under what conditions does the reverse hold, i.e., when is it possible to attain any point in the interior of \mathcal{F}^* as PPE payoffs?

In order to attain nearly efficient payoffs, players need to be able to support cooperation by detecting and appropriately punishing deviations. In PPEs, where strategies are public, all such punishment should occur solely based on the public signals. As a result, the public signals should

be distributed such that they allow players to statistically distinguish between deviations by two different players, as well as different deviations by the same player. We now formally specify these conditions. The first condition, referred to as *individual full rank*, gives a sufficient condition under which deviations by a single player are statistically distinguishable; i.e., the distribution over signals induced by some profile $\boldsymbol{\rho}$ are different from that induced by any $(\rho'_i, \boldsymbol{\rho}_{-i})$ for $\rho'_i \neq \rho_i$. Formally,

DEFINITION 1. The profile $\boldsymbol{\rho}$ has individual full rank for player i if given the strategies of the other players, $\boldsymbol{\rho}_{-i}$, the $|R_i| \times |B|$ matrix $A_i(\boldsymbol{\rho}_{-i})$ with entries $[A_i(\boldsymbol{\rho}_{-i})]_{r_i, b} = \pi(b|r_i, \boldsymbol{\rho}_{-i})$ has full row rank. That is, the $|R_i|$ vectors $\{\pi(\cdot|r_i, \boldsymbol{\rho}_{-i})\}_{r_i \in R_i}$ are linearly independent.

The second general condition, *pairwise full rank*, is a strengthening of individual full rank to pairs of players. In essence, it ensures that deviations by players i and j are distinct, as they introduce different distributions over public outcomes. Formally,

DEFINITION 2. The profile $\boldsymbol{\rho}$ has pairwise full rank for players i and j if the $(|R_i| + |R_j|) \times |B|$ matrix $A_{ij}(\boldsymbol{\rho}) := [A_i(\boldsymbol{\rho}_{-i}); A_j(\boldsymbol{\rho}_{-j})]$ has rank $|R_i| + |R_j| - 1$.

Therefore, given an adequate public monitoring signal, we have the following folk theorem under imperfect public monitoring.

THEOREM 1 (The imperfect public monitoring folk theorem, Fudenberg et al. (1994)). Assume R is finite, the set of feasible payoffs $\mathcal{F}^\dagger \subset \mathbb{R}^N$ has non-empty interior, and all the pure action equilibria leading the extreme points of \mathcal{F}^\dagger have pairwise full rank for all pairs of players. If the minmax payoff profile $\underline{\mathbf{v}} = (\underline{v}_1, \dots, \underline{v}_N)$ is inefficient, and the minmax profile $\hat{\boldsymbol{\rho}}^i$ has individual full rank for each player i , then for any profile of payoffs $\mathbf{v} \in \text{int}\mathcal{F}^*$, there exists a discount factor $\underline{\delta} < 1$, such that for all $\delta \in (\underline{\delta}, 1)$, $\mathbf{v} \in \mathcal{E}(\delta)$.

3.2. Supporting Cooperation in Information Sharing with Public Monitoring

We now verify that the above folk theorem applies to information sharing games with imperfect public monitoring structure given by (2). That is, when the firms are sufficiently patient, they can sustain cooperation on full security information sharing in a repeated setting, by making their disclosure decisions based only on the imperfect, publicly announced observations of the monitor about their past actions. To this end, we need to verify that the conditions of the folk theorem, in particular those on the informativeness of the public monitoring signal, hold for (2). First, note that the public signal b has 2^N possible outcomes; we view each signal as a binary string and assume the columns of the following matrices are ordered according to the decimal equivalent of these binary strings.

We first verify that the minmax profile of the repeated information sharing game has individual full rank for any firm. The minmax action profile for some firm i , $\hat{\mathbf{r}}^i$, is all firms concealing their

information, i.e., $\hat{\mathbf{r}}^i = \mathbf{0}$. Consider deviations by firm 1 (by the symmetric nature of the game, the same argument holds for other firms). Then $A_1(\hat{\mathbf{r}}^i)$ is given by:

$$\mathbf{b} = \begin{matrix} & (0,0,\dots,0) & (1,0,\dots,0) & \dots & (0,1,\dots,1) & (1,1,\dots,1) \\ r_1=0 & \left(\begin{array}{ccccc} \alpha^N & (1-\alpha)\alpha^{N-1} & \dots & \alpha(1-\alpha)^{N-1} & (1-\alpha)^N \\ \epsilon\alpha^{N-1} & (1-\epsilon)\alpha^{N-1} & \dots & \epsilon(1-\alpha)^{N-1} & (1-\epsilon)(1-\alpha)^{N-1} \end{array} \right) \end{matrix}$$

The rows of the above matrix are linearly independent (given $\alpha \neq \epsilon$), and hence the minmax profiles have individual full rank for all firms.

We also need to verify that all pure strategy action profiles have pairwise full rank. We do so for $\mathbf{r}_k := (1,1,\dots,1,0,0,\dots,0)$, where the first k firms disclose, and the remainder $N-k$ conceal; other profiles can be shown similarly. For the profile \mathbf{r}_k , first consider the firms $i=1$ and $j=N$. The matrix $A_{ij}(\mathbf{r}_k)$ is given by:

$$\mathbf{b} = \begin{matrix} & (0,0,\dots,0) & (1,0,\dots,0) & \dots & (0,1,\dots,1) & (1,1,\dots,1) \\ r_1=0 & \left(\begin{array}{ccccc} \alpha \cdot \epsilon^{k-1} \cdot \alpha^{N-k-1} \cdot \alpha & (1-\alpha) \cdot \epsilon^{k-1} \cdot \alpha^{N-k-1} \cdot \alpha & \dots & \alpha \cdot (1-\epsilon)^{k-1} \cdot (1-\alpha)^{N-k} & (1-\alpha) \cdot (1-\epsilon)^{k-1} \cdot (1-\alpha)^{N-k} \\ \epsilon \cdot \epsilon^{k-1} \cdot \alpha^{N-k-1} \cdot \alpha & (1-\epsilon) \cdot \epsilon^{k-1} \cdot \alpha^{N-k-1} \cdot \alpha & \dots & \epsilon \cdot (1-\epsilon)^{k-1} \cdot (1-\alpha)^{N-k} & (1-\epsilon)^k \cdot (1-\alpha)^{N-k} \\ r_N=0 & \left(\begin{array}{ccccc} \epsilon \cdot \epsilon^{k-1} \cdot \alpha^{N-k-1} \cdot \alpha & (1-\epsilon) \cdot \epsilon^{k-1} \cdot \alpha^{N-k-1} \cdot \alpha & \dots & \epsilon \cdot (1-\epsilon)^{k-1} \cdot (1-\alpha)^{N-k} & (1-\epsilon)^k \cdot (1-\alpha)^{N-k} \\ r_N=1 & \left(\begin{array}{ccccc} \epsilon \cdot \epsilon^{k-1} \cdot \alpha^{N-k-1} \cdot \epsilon & (1-\epsilon) \cdot \epsilon^{k-1} \cdot \alpha^{N-k-1} \cdot \epsilon & \dots & \epsilon \cdot (1-\epsilon)^{k-1} \cdot (1-\alpha)^{N-k-1} \cdot (1-\epsilon) & (1-\epsilon)^k \cdot (1-\alpha)^{N-k-1} \cdot (1-\epsilon) \end{array} \right) \end{array} \right) \end{matrix}$$

Note that the rows corresponding to $r_1=1$ and $r_N=0$ are the same: indeed when both firms follow the prescribed strategy, the distribution of the signals is consistent. It is straightforward to verify that the above has row rank 3; i.e., removing the common row, the three remaining rows are linearly independent. As a result, \mathbf{r}_k has pairwise full rank for firms $i=1$ and $j=N$. A similar procedure follows for other pairs of firms i,j , the remaining pure action profiles, verifying that all have pairwise full rank.

We have therefore established the following proposition:

PROPOSITION 1. *The conditions of the folk theorem of Section 3.1 hold with the public signals distributed according to (2). As a result, when the firms are sufficiently patient; i.e., they value the future outcomes of their information sharing agreement, it is possible for them to nearly efficiently cooperate on full information disclosure through repeated interactions.*

3.3. Constructing Public Strategies: An Example

In this section, we present the process through which equilibrium public strategies leading to a desired payoff profile are constructed. To simplify the illustration, we consider a two player prisoner's dilemma game with payoff matrix given by Table 1.

We first present an overview of the idea behind constructing the equilibrium strategies. The utility of firms at each step of the game can be decomposed into their current payoff, plus the continuation payoff; i.e., the expected payoff for the remainder of the game depending on the observed public monitoring output. Therefore, to achieve an average payoff profile \mathbf{v} as equilibrium

	C	D
C	$G - L, G - L$	$-L, G$
D	$G, -L$	$0, 0$

Table 1 Firms' payoffs in a two-person prisoner's dilemma game

in the repeated game, the action profile and the continuation payoffs should be selected so as to maximize firms' expected payoff.

Formally, we say \mathbf{v} is decomposed by \mathbf{r} on a set W using a mapping $\gamma : B \rightarrow W$ if:

$$\begin{aligned} v_i &= (1 - \delta)u_i(\mathbf{r}) + \delta E[\gamma_i(b)|\mathbf{r}] \\ &\geq (1 - \delta)u_i(r'_i, \mathbf{r}_{-i}) + \delta E[\gamma_i(b)|r'_i, \mathbf{r}_{-i}], \quad \forall r'_i \in R_i, \forall i. \end{aligned} \quad (4)$$

Here, the mapping γ determines firms' continuation payoffs (selected from a set W) following each signal $b \in B$. The goal is thus to set $W = \mathcal{E}(\delta)$ (the set of PPE payoffs), and find appropriate actions \mathbf{r} and mappings γ decomposing (i.e., satisfying (4) for) payoff profiles $\mathbf{v} \in \mathcal{E}(\delta)$. We can then conclude that any payoff profile \mathbf{v} for which the above decomposition is possible, will be attainable as a PPE average payoff, as we can recursively decompose the selected continuation payoffs on $\mathcal{E}(\delta)$ as well. This procedure thus characterizes the set of payoffs that can be attained using public strategies.

However, the set of decomposable payoffs on arbitrary sets W is in general hard to characterize; let's instead consider the simpler decomposition on half-spaces $H(\lambda, \lambda \cdot \mathbf{v}) := \{\mathbf{v}' \in \mathbb{R}^N : \lambda \cdot \mathbf{v}' \leq \lambda \cdot \mathbf{v}\}$. With $W = H(\lambda, \lambda \cdot \mathbf{v})$, (4) can be re-written as:

$$\begin{aligned} v_i &= u_i(\mathbf{r}) + E[\bar{\gamma}_i(b)|\mathbf{r}] \geq u_i(r'_i, \mathbf{r}_{-i}) + E[\bar{\gamma}_i(b)|r'_i, \mathbf{r}_{-i}], \quad \forall r'_i \in R_i, \forall i, \\ \text{and, } \lambda \cdot \bar{\gamma}(b) &\leq 0, \quad \forall b \in B, \end{aligned} \quad (5)$$

where $\bar{\gamma} : B \rightarrow \mathbb{R}^N$, and $\bar{\gamma}_i(b) = \frac{\delta}{1-\delta}(\gamma_i(b) - v_i)$. We refer to x as the normalized continuation payoffs.

It can be shown (see Mailath and Samuelson (2006)) that characterizing the set of attainable PPE payoffs $\mathcal{E}(\delta)$ is equivalent to finding the maximum average payoffs that can be decomposed on half-spaces using different actions \mathbf{r} and in various directions λ . We therefore first find the maximum average payoffs \mathbf{v} enforceable on half-spaces (i.e., satisfying (5), and with $\lambda \cdot \bar{\gamma}(b) = 0$ whenever possible), for each action profile \mathbf{r} and direction λ . We will then select the best action \mathbf{r} for each direction, and finally take the intersection over all possible directions λ to characterize $\mathcal{E}(\delta)$.⁵

⁵ Define $k^*(\lambda; \mathbf{r}) := \lambda \cdot \bar{\mathbf{v}}$, where $\bar{\mathbf{v}}$ is the maximum payoff profile satisfying (5). It can be shown that $k^*(\lambda; \mathbf{r}) \leq \lambda \cdot u(\mathbf{r})$, and so the maximum is attained when \mathbf{r} is *orthogonally enforced* (whenever possible); i.e., $\lambda \cdot \bar{\gamma}(b) = 0$ in (5). Let $k^*(\lambda) = \sup_{\mathbf{r}} k^*(\lambda; \mathbf{r})$. Intuitively, $k^*(\lambda)$ is a bound on the average payoff for firms for which the incentive constraints are satisfied. Let $H^*(\lambda) := H(\lambda, k^*(\lambda))$ be the corresponding *maximal half-space*. Then, that the set of PPE payoffs is contained in the intersection of these maximal half-spaces; i.e., $\mathcal{E}(\delta) \subseteq \cap_{\lambda} H^*(\lambda) := \mathcal{M}$, and that the reverse is also true for sufficiently large δ ; i.e., $\lim_{\delta \rightarrow 1} \mathcal{E}(\delta) = \mathcal{M}$. We refer the interested reader to Mailath and Samuelson (2006) for more details.

We now find the average payoffs decomposable on half-spaces for the prisoner's dilemma game in Table 1. Let us first consider profile $\mathbf{r} = (1, 1)$,⁶ and an arbitrary direction $\lambda = (\lambda_1, \lambda_2)$. Setting $\lambda \cdot \bar{\gamma}(b) = 0$, (5) reduces to:

$$\begin{aligned}
G - L &= G - L + (\epsilon^2 \bar{\gamma}_1(0, 0) + \epsilon(1 - \epsilon) \bar{\gamma}_1(0, 1) + (1 - \epsilon) \epsilon \bar{\gamma}_1(1, 0) + (1 - \epsilon)^2 \bar{\gamma}_1(1, 1)) \\
&\geq G + (\epsilon \alpha \bar{\gamma}_1(0, 0) + \alpha(1 - \epsilon) \bar{\gamma}_1(0, 1) + (1 - \alpha) \epsilon \bar{\gamma}_1(1, 0) + (1 - \epsilon)(1 - \alpha) \bar{\gamma}_1(1, 1)) \\
&\text{and ,} \\
G - L &= G - L + (\epsilon^2 \bar{\gamma}_2(0, 0) + \epsilon(1 - \epsilon) \bar{\gamma}_2(0, 1) + (1 - \epsilon) \epsilon \bar{\gamma}_2(1, 0) + (1 - \epsilon)^2 \bar{\gamma}_2(1, 1)) \\
&\geq G + (\epsilon \alpha \bar{\gamma}_2(0, 0) + \epsilon(1 - \alpha) \bar{\gamma}_2(0, 1) + (1 - \epsilon) \alpha \bar{\gamma}_2(1, 0) + (1 - \epsilon)(1 - \alpha) \bar{\gamma}_2(1, 1)) \\
&\text{and ,} \\
\lambda_1 \bar{\gamma}_1(b) + \lambda_2 \bar{\gamma}_2(b) &= 0, \quad \forall b \in B .
\end{aligned}$$

Substituting for $\bar{\gamma}_2(b)$ using the last equation, and writing the inequalities as equalities, finding the normalized continuation payoffs is equivalent to solving the following system of equations:

$$\begin{pmatrix} \epsilon^2 & \epsilon(1 - \epsilon) & (1 - \epsilon)\epsilon & (1 - \epsilon)^2 \\ \alpha\epsilon & \alpha(1 - \epsilon) & (1 - \alpha)\epsilon & (1 - \alpha)(1 - \epsilon) \\ \epsilon^2 & \epsilon(1 - \epsilon) & (1 - \epsilon)\epsilon & (1 - \epsilon)^2 \\ \alpha\epsilon & (1 - \alpha)\epsilon & \alpha(1 - \epsilon) & (1 - \alpha)(1 - \epsilon) \end{pmatrix} \begin{pmatrix} \bar{\gamma}_1(0, 0) \\ \bar{\gamma}_1(0, 1) \\ \bar{\gamma}_1(1, 0) \\ \bar{\gamma}_1(1, 1) \end{pmatrix} = \begin{pmatrix} 0 \\ -L \\ 0 \\ L \frac{\lambda_2}{\lambda_1} \end{pmatrix} .$$

The first and third rows represent the same equations (corresponding to the equilibrium outcome). Removing the third row and performing row-reduction on the remaining matrix, the continuation payoffs should satisfy the following set of equations:

$$\begin{aligned}
\epsilon \bar{\gamma}_1(0, 0) + (1 - \epsilon) \bar{\gamma}_1(0, 1) &= \frac{-L}{\alpha\kappa} \frac{1 - \epsilon}{\epsilon} \\
\epsilon \bar{\gamma}_1(1, 0) + (1 - \epsilon) \bar{\gamma}_1(1, 1) &= \frac{L}{\alpha\kappa} \\
-\bar{\gamma}_1(0, 1) + \bar{\gamma}_1(1, 0) &= \frac{L}{\epsilon\alpha\kappa} \left(\frac{\lambda_2}{\lambda_1} + 1 \right) ,
\end{aligned}$$

where $\kappa := \frac{1 - \epsilon}{\epsilon} - \frac{1 - \alpha}{\alpha} > 0$. The above is an underdetermined system, and thus has infinitely many solutions depending on the designer's choice of continuation payoffs. We construct and interpret one such possibility.

Let's set $\bar{\gamma}_1(1, 1) = 0$, implying $\bar{\gamma}_2(1, 1) = 0$ as well. This means if the signal indicates that both firms are cooperating with high probability, there is no need for punishments, so both firms expect their continuation payoff to remain unchanged (i.e., equal to their current payoff). Given this choice, we can solve for the remaining normalized continuation payoffs, illustrated in Table 2.

⁶ Note that decomposing using $(0, 0)$ is not considered as it leads to the maximal half-space \mathbb{R}^2 . It thus provides no information on the set of attainable payoffs as we already know that $\mathcal{E}(\delta) \subseteq \mathbb{R}^2$.

	$\bar{\gamma}_1(b)$	$\bar{\gamma}_2(b)$
$b=(0,0)$	$\frac{L}{\epsilon\alpha\kappa} \frac{1-\epsilon}{\epsilon} \left(\frac{\lambda_2}{\lambda_1} - 1 \right)$	$\frac{L}{\epsilon\alpha\kappa} \frac{1-\epsilon}{\epsilon} \left(\frac{\lambda_1}{\lambda_2} - 1 \right)$
$b=(0,1)$	$-\frac{\lambda_2}{\lambda_1} \frac{L}{\epsilon\alpha\kappa}$	$\frac{L}{\epsilon\alpha\kappa}$
$b=(1,0)$	$\frac{L}{\epsilon\alpha\kappa}$	$-\frac{\lambda_1}{\lambda_2} \frac{L}{\epsilon\alpha\kappa}$
$b=(1,1)$	0	0

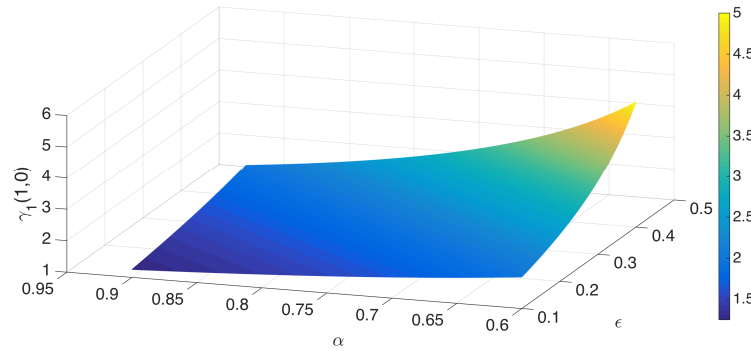
Table 2 An example of normalized continuation payoff choices.

Intuition. These normalized continuation payoffs can be intuitively interpreted as follows. Fix a direction with $\lambda_1, \lambda_2 > 0$ (similar interpretations follow for other directions). Then, given a signal $b = (1, 0)$, which is more likely under a deviation by firm 2, firm 1 expects a higher continuation payoff ($\bar{\gamma}_1(1, 0) > 0$), while the suspect deviator expects a lower one ($\bar{\gamma}_2(1, 0) < 0$).⁷ A similar intuition applies to the continuations under the signal $(0, 1)$. On the other hand, with $b = (0, 0)$, either firm 1 or 2 will be punished, depending on the direction λ . Specifically, for a direction $\lambda_1 = \lambda_2$, neither firm expects a change in her continuation payoff. Note that with $\lambda_1 = \lambda_2$, the change in continuation payoffs between the outcomes $(0, 1)$ and $(1, 0)$, as well as among firms in either outcome, are also of equal size. Note also that both firms are never punished simultaneously under any outcome, so as to maintain a high average payoff.

Finally, it is worth noting the effect of the monitoring accuracy, α and ϵ , on the normalized continuation payoffs. Consider direction $\lambda_1 = \lambda_2 = 1$, and fix $L = 1$. First, note that $\epsilon\alpha\kappa = \alpha(1 - \epsilon) - \epsilon(1 - \alpha)$ is increasing in α and decreasing in ϵ . This is illustrated in Fig. 1, which shows the dependence of $\bar{\gamma}_1(1, 0)$ on the monitoring parameters. As a result, as the monitoring technology becomes more accurate, i.e., α increases and/or ϵ decreases, the size of the normalized continuation payoffs for firms, when $(1, 0)$ or $(0, 1)$ is observed, becomes smaller. This is because, as monitoring becomes accurate, signals indicating deviations (despite equilibrium being played) are more likely to be due to monitoring errors rather than actual deviations, and therefore the required continuation punishments/rewards become less severe to maintain firms' average payoff high.

We conclude that in general, using the described procedure, we can decompose payoff profiles in the half-spaces $H(\lambda, k^*(\lambda, (1, 1)))$, where $k^*(\lambda, (1, 1)) = \lambda \cdot u(1, 1) = (G - L)(\lambda_2 + \lambda_1)$, using the action profile $\mathbf{r} = (1, 1)$ and continuation payoffs determined as above. Using a similar procedure, the corresponding half-spaces for the remaining action profiles will have $k^*(\lambda, (0, 1)) = G\lambda_1 - L\lambda_2$ and $k^*(\lambda, (1, 0)) = G\lambda_2 - L\lambda_1$.

⁷ It is worth emphasizing that due to the equilibrium construction, firms are both playing $r_i = 1$; nevertheless, punishments on the equilibrium path happen due to the imperfection of monitoring.

Figure 1 $\tilde{\gamma}_1(1,0)$

We next choose, for a given direction λ , the action for which the corresponding half-spaces covers a larger set of average payoffs; i.e. $k^*(\lambda) = \max_{\mathbf{r}} \{G\lambda_2 - L\lambda_1, G\lambda_1 - L\lambda_2, (G-L)(\lambda_1 + \lambda_2)\}$; which leads to:

$$k^*(\lambda) = \begin{cases} G\lambda_2 - L\lambda_1 & \lambda_2 \geq \frac{G}{L}\lambda_1 \\ (G-L)(\lambda_1 + \lambda_2) & \frac{L}{G}\lambda_1 \leq \lambda_2 \leq \frac{G}{L}\lambda_1 \\ G\lambda_1 - L\lambda_2 & \lambda_1 \geq \frac{G}{L}\lambda_2 \end{cases}$$

Finally, it is straightforward to show that the intersection of half-spaces $H(\lambda, k^*(\lambda))$, as λ ranges over \mathbb{R}^2 , is equivalent to the set of feasible and strictly individually rational payoffs of the two-person prisoner's dilemma game of Table 1. That is, it is possible to find an action profile \mathbf{r} and the corresponding continuation payoff mapping γ (constructed as described above), so as to incentivize any feasible strictly individually rational payoff profile.

4. Imperfect Private Monitoring: The Role of Communication

In this section, we consider the use of private monitoring in providing inter-temporal incentives for information sharing. Unlike repeated games with imperfect public monitoring, relatively less is known about games with private monitoring (Kandori (2002)).

In particular, we are interested in a folk theorem for this repeated game; i.e., a full characterization of payoffs that can be achieved when firms only have private observations, if firms are sufficiently patient. As discussed in Section 3, with imperfect public monitoring, Fudenberg et al. (1994) present a folk theorem under relatively general conditions. The possibility of this result hinges heavily on that firms share common information on each others' actions (i.e., the public monitoring outcome), as a result of which it is possible to recover a recursive structure for the game, upon which the folk theorem is based. However, a similar folk theorem with private monitoring remained an open problem until recently,⁸ mainly due to the lack of a common public signal.

⁸ A recent advance in the field is by Sugaya (2011, 2013), who presents a folk theorem for repeated games with imperfect private monitoring, without requiring cheap talk communication or public randomization. We however analyze the application of a folk theorem using communication, to draw a closer parallel with the public monitoring structure of the previous section.

Nevertheless, the possibility of cooperation, and in particular folk theorems, have been shown to exist for some particular classes of these games. Examples include:

- Games in which firms are allowed to communicate (cheap talk) after each period. This approach has been proposed in Compte (1998), Kandori and Matsushima (1998), and in essence, uses the signals collected through communication as a public signal, allowing participants to coordinate on cooperation.
- Games in which firms have public actions (e.g., announcement of sanctions) in addition to private decisions (here, disclosure decisions), as proposed by Park (2011) for the study of international trade agreements. Intuitively, public actions serve a similar purpose as communication, allowing participants to signal the initiation of punishment phases.
- Games with almost public monitoring, i.e., private monitoring with signals that are sufficiently correlated. With such signals, Mailath and Morris (2002) proves a folk theorem for almost-perfect and almost-public monitoring.

In this section, we present the folk theorem with private monitoring and communication due to Kandori and Matsushima (1998) in Section 4.1, and in Section 4.2, verify that it applies to NPD information sharing games with monitoring given by (1).

4.1. The Folk Theorem with Imperfect Private Monitoring and Communication

In this section, we describe the folk theorem with imperfect private monitoring, allowing for communication between players, due to Kandori and Matsushima (1998).

The stage game is similar to the setup of Fudenberg et al. (1994) in Section 3.1. Consider N rational players. At the stage game, each player i chooses an action $r_i \in R_i$. Let $\mathbf{r} \in R := \prod_{i=1}^N R_i$ denote a profile of actions. At the end of each stage, each player privately observes an outcome $b_i \in B_i$, where B_i is a finite set of possible signals. The probability of observing the profile of private signals $\mathbf{b} \in B := \prod_{i=1}^N B_i$ following \mathbf{r} is given by the joint distribution $\pi(\mathbf{b}|\mathbf{r})$. Assume π has full support; i.e., $\pi(\mathbf{b}|\mathbf{r}) > 0$, $\forall \mathbf{b}, \forall \mathbf{r}$. Let $u_i^*(r_i, b_i)$ be the utility of player i when she plays r_i and observes the signal b_i . Note that i 's utility depends on others' actions only through b_i , and thus the stage payoffs are not informative about others' actions. The expected stage game payoff for user i when \mathbf{r} is played is therefore given by:

$$u_i(\mathbf{r}) = \sum_{\mathbf{b} \in B} u_i^*(r_i, b_i) \pi(\mathbf{b}|\mathbf{r}). \quad (6)$$

The definition of the minmax action profiles $\boldsymbol{\rho}^i$ and the set of feasible and strictly individually rational payoffs $\mathcal{F}^* \subset \mathbb{R}^N$ is the same as those in Section 3.1. However, in addition to the private nature of signals b_i , the current model differs from the setup in Section 3.1 in that we allow the players to communicate in this game. Formally, after choosing the action r_i and observing the

signal b_i , each player i will publicly announce a message $m_i \in M_i$, selected from the finite set of possible messages M_i . Let $M = \prod_{i=1}^N M_i$.

Consequently, the strategy $s_i = (r_i, m_i)$ of each player consists of both an action r_i and a message m_i . In particular, when the game is played repeatedly often, the strategy specifies a choice for each time step t ; i.e., $r_i = (r_i(t))_{t=0}^\infty$ and $m_i = (m_i(t))_{t=0}^\infty$, where:

$$\begin{aligned} r_i(t) &: R_i^{t-1} \times B_i^{t-1} \times M^{t-1} \rightarrow \Delta(R_i) , \\ m_i(t) &: R_i^t \times B_i^t \times M^{t-1} \rightarrow \Delta(M_i) . \end{aligned}$$

Let $r_i^t = (r_i(0), \dots, r_i(t))$. Define b_i^t and m^t similarly. Then, the private history of player i at the end of time t is given by $h_i^t := (r_i^t, b_i^t)$, and the public history is $h^t := m^t$. Therefore, players' strategies depend on both private and public histories. Given the strategy profiles $\mathbf{s} = (s_1, \dots, s_N)$, and assuming that players discount future payoffs by a discount factor δ , a player's average payoff throughout the repeated game is given by $(1 - \delta) \sum_{t=0}^\infty \delta^t u_i(\mathbf{s}(t))$. Each player i is choosing her strategy s_i to maximize the expected value of this expression.

We are interested in characterizing the payoffs attainable by the strategy profiles \mathbf{s} that are a *sequential equilibrium* of the game. Formally, \mathbf{s} is a sequential equilibrium of the game if for every player and her history (h_i^t, h^t) , $s_i|_{(h_i^t, h^t)}$ is a best reply to $E[s_{-i}|_{h_{-i}^t, h^t}|h_i^t]$. That is, a player is best-responding according to her belief over private histories of other players, in particular those which are consistent with her own private history. Let $V(\delta)$ denote the set of sequential equilibrium average payoffs when the discount factor is δ . We are interested in identifying conditions under which $V(\delta) \subseteq \mathcal{F}^*$.

Recall that in the game of imperfect public monitoring, conditioning of strategies on the publicly observed signal allowed players to coordinate, and to recover a recursive structure in the game. The possibility of communication allows for recovering a similar recursive structure in games with private monitoring. The equilibrium strategies leading to nearly efficient payoffs will be constructed as follows. At the end of each period t , each player i is asked to report her privately observed signal as her message; i.e., $m_i(t) = b_i(t)$. To make sure that players truthfully report their signals, the equilibrium strategies use this private information to determine *other* players' deviations and future payoffs, and maintain i 's payoff independent of her report. As a result, truthful reporting of privately observed signals will be a (weak) best-response.^{9,10} It remains to ensure, following a

⁹ It is also possible to make truth reporting a *strict* best-response if players' privately observed signals are mutually correlated; see (Kandori and Matsushima 1998, Section 4.2).

¹⁰ Note that unlike Section 3, each player will be playing a private strategy at equilibrium, as she is using her private information as her message m_i . However, the choice of action r_i will be based only on the public information; i.e., the disclosed messages available to all players.

rationale similar to the folk theorem of Section 3.1, that the available signals are informative enough, in the sense that they allow players to distinguish between different deviations of individuals, and to differentiate among deviations by different players.

The required conditions on the informativeness of the players' signals are as follows. First, define the following vectors:

$$\begin{aligned} p_{-i}(\mathbf{r}) &:= (\pi_{-i}(\mathbf{b}_{-i}|\mathbf{r}))_{\mathbf{b}_{-i} \in B_{-i}} , \\ p_{-ij}(\mathbf{r}) &:= (\pi_{-ij}(\mathbf{b}_{-ij}|\mathbf{r}))_{\mathbf{b}_{-ij} \in B_{-ij}} \\ Q_{ij}(\mathbf{r}) &:= \{p_{-ij}(\mathbf{r}_{-i}, r'_i) | r'_i \in R_i \setminus \{r_i\}\} , \end{aligned}$$

where $B_{-i} := \prod_{k \neq i} B_k$, $B_{-ij} := \prod_{k \neq i, j} B_k$, and π_{-i} and π_{-ij} are marginal distributions of the joint distribution $\pi(\mathbf{b}|\mathbf{r})$ of privately observed signals. The three sufficient conditions on signals can be expressed accordingly.

Condition 1 *At the minmax strategy profile of a player i , $\hat{\rho}^i$, for any player $j \neq i$ and any mixed strategy $\rho'_j \in \Delta(R_j)$, either*

$$\begin{aligned} (i) \quad & p_{-j}(\hat{\rho}^i) \neq p_{-j}(\hat{\rho}_{-j}^i, \rho'_j) \quad \text{or,} \\ (ii) \quad & p_{-j}(\hat{\rho}^i) = p_{-j}(\hat{\rho}_{-j}^i, \rho'_j) \quad \text{and} \quad u_j(\hat{\rho}^i) \geq u_j(\hat{\rho}_{-j}^i, \rho'_j). \end{aligned}$$

Condition (C1) states that at the minmax profile of any player, a deviation by another player is either statistically distinguishable, and if not, it reduces the payoff of the deviator, and is hence not profitable. This assumption ensures that we can provide incentives to players to punish (minmax) one another. Note that this requirement is similar to (but weaker than) the individual full rank condition in the folk theorem of Section 3.1.

Condition 2 *For each pair of players $i \neq j$, and each pure action equilibrium \mathbf{r} leading to an extreme point of the payoff set \mathcal{F}^\dagger , we have:*

$$p_{-ij}(\mathbf{r}) \notin \text{co}(Q_{ij}(\mathbf{r}) \cap Q_{ji}(\mathbf{r})) ,$$

where $\text{co}(X)$ denotes the convex hull of the set X .

Recall that $Q_{ij}(\mathbf{r})$ denotes the vector of distribution of beliefs of players other than i and j , when player i is deviating. (C2) therefore requires that a deviation by either i or j (but not both) is statistically detected by the remaining players.

Condition 3 *For each pair of players $i \neq j$, and each pure action equilibrium \mathbf{r} leading to an extreme point of the payoff set \mathcal{F}^\dagger , we have:*

$$\text{co}(Q_{ij}(\mathbf{r}) \cup p_{-ij}(\mathbf{r})) \cap \text{co}(Q_{ji}(\mathbf{r}) \cup p_{-ji}(\mathbf{r})) = \{p_{-ij}(\mathbf{r})\} .$$

Finally, (C3) requires that players other than i, j can statistically distinguish deviations by i from deviations by j , as the resulting distribution on \mathbf{b}_{-ij} will be different under either player's deviation. In other words, the only consistent distribution arises when neither player is deviating. It is worth mentioning that (C2) and (C3) hold when the pairwise full rank condition of the folk theorem in Section 3.1 holds.

Therefore, given adequate private monitoring signals and communication, we have the following folk theorem under imperfect private monitoring.

THEOREM 2 (The Imperfect private monitoring with communication folk theorem). *[Kandori and Matsushima (1998)] Assume that there are more than two players ($N > 2$), and the set of feasible and strictly individual rational payoffs $\mathcal{F}^* \subset \mathbb{R}^N$ has non-empty interior (and therefore dimension N). Then, if the monitoring of players satisfy conditions (C1), (C2), and (C3), any interior payoff profile $\mathbf{v} \in \text{int}\mathcal{F}^*$ can be achieved as a sequential equilibrium average payoff profile of the repeated game with communication, when δ is close enough to 1.*

4.2. Supporting Cooperation in Information Sharing with Private Monitoring and Communication

We now verify that the above folk theorem applies to information sharing games with imperfect private monitoring structure given by (1). That is, when the firms are sufficiently patient, they can sustain cooperation on full security information sharing in a repeated setting, by truthfully revealing their private signals, and making their disclosure decisions based only on the imperfect, publicly announced collective observation about their past actions. To this end, we need to verify that the three conditions of the folk theorem on the informativeness of the monitoring signal hold for the joint distribution of the private signals in (1). However, note that once these signals are truthfully reported, it is as if we have access to $N - 1$ independent realizations of the public monitoring distribution in (2). Assume that to test the conditions of the folk theorem, we randomly choose one of the available cross-observations about a possible deviator (from all players other than the suspect for (C1), or other than the two suspects for (C2) and (C3)) and test the statistical distinguishability of the signal. With this method, the joint distribution of the private signal that is being tested will be equivalent to the public monitoring distribution of (2).

On the other hand, as mentioned in Section 4.1, the conditions of the folk theorem for imperfect private monitoring are weaker than the full rank conditions required by the folk theorem with imperfect public monitoring: (C1) is a weaker condition than individual full rank, and (C2) and (C3) are weaker versions of pairwise full rank. This is because the individual and pairwise full rank conditions require linear independence of their corresponding signal distributions, while (C1)-(C3) are stated in terms of convex combinations. As shown in Section 3.2, the distribution in (2) satisfies

the full rank conditions, and consequently, the joint private distribution resulting from (1) satisfies (C1)-(C3). We have therefore established the following proposition.

PROPOSITION 2. *The conditions of the folk theorem of Section 4.1 hold with private monitoring (1) and communication. As a result, when the firms are sufficiently patient, i.e., they value the future outcomes of their information sharing agreement, and are allowed to communicate their private signals, it is possible for them to nearly efficiently cooperate on full information disclosure through repeated interactions.*

5. Conclusion

We modeled information sharing agreements among firms as an N-person prisoner's dilemma game, and equipped it with a simple binary monitoring structure. We proposed a repeated-game approach to this problem, and discussed the role of monitoring (private vs. public) on determining whether inter-temporal incentives can lead to the support of cooperation (i.e., full disclosure). Specifically, we showed that a rating/monitoring system can play a crucial role in providing a common public signal which, despite being imperfect, can be used to design inter-temporal incentives that lead firms to cooperate on information sharing. We also showed that in the absence of a monitor, if the firms are provided with a platform to communicate their privately observed beliefs on each others' adherence to the agreement, it is again possible to design similar inter-temporal incentives.

An important requirement for the folk theorem, and consequently the design of inter-temporal incentives, is to ensure that firms are sufficiently patient (i.e., they place significant value on their future interactions), as characterized by having discount factors higher than $\underline{\delta}$. Despite the fact that the proposed binary monitoring structures in (1) and (2) are informative enough for the folk theorem to hold, their accuracy, (α, ϵ) , will impact the requirement on firms' patience, $\underline{\delta}$. Characterizing the dependence of $\underline{\delta}$ on (α, ϵ) is a main direction of future work. Particularly, we have only considered one method for using firms' communication of their privately observed signals to establish the folk theorem of Section 4.1. Determining the optimal method for combining firms' inputs to ensure the lowest $\underline{\delta}$ remains an interesting question.

Another possible direction is to consider the design of inter-temporal incentives when both types of public and private monitoring are available. It is indeed still possible to have firms coordinate based on the public monitoring system's report alone (i.e., use public strategies); nevertheless, it may also be possible to employ *private strategies*, in which firms use both their own observations, as well as the public signal. Private strategies may lead to higher payoffs than those attainable through public strategies alone (Mailath and Samuelson 2006, Chapter 10), thus making their study of interest to either lower the required discount factor, or when the monitoring signals are not informative enough for a public monitoring folk theorem to hold.

Finally, we have assumed that the monitoring, as well as its accuracy, are fixed and available to firms at no additional cost. Analyzing the effects of costly monitoring on firms' incentives is another direction of future work.

Acknowledgments

This material is based on research sponsored by the Department of Homeland Security (DHS) Science and Technology Directorate, Homeland Security Advanced Research Projects Agency (HSARPA), Cyber Security Division (DHS S&T/HSARPA/CSD), BAA 11-02 via contract number HSHQDC-13-C-B0015.

References

- Bonacich, Phillip, Gerald H Shure, James P Kahan, Robert J Meeker. 1976. Cooperation and group size in the n-person prisoners' dilemma. *Journal of Conflict Resolution* **20**(4) 687–706.
- Campbell, Katherine, Lawrence A Gordon, Martin P Loeb, Lei Zhou. 2003. The economic cost of publicly announced information security breaches: empirical evidence from the stock market. *Journal of Computer Security* **11**(3) 431–448.
- Cavusoglu, Huseyin, Birendra Mishra, Srinivasan Raghunathan. 2004. The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers. *International Journal of Electronic Commerce* **9**(1) 70–104.
- Claburn, Thomas. 2008. Data breaches made possible by incompetence, carelessness. URL <http://www.darkreading.com/risk-management/data-breaches-made-possible-by-incompetence-carelessness> Retrieved on 2016-02-24.
- Compte, Olivier. 1998. Communication in repeated games with imperfect private monitoring. *Econometrica* 597–626.
- DHS. 2015a. Enhancing resilience through cyber incident data sharing and analysis. URL <https://www.dhs.gov/sites/default/files/publications/Data%20Categories%20White%20Paper%20-%20508%20> Retrieved on 2016-02-24.
- DHS. 2015b. Information sharing and analysis organizations (ISAOs). URL <http://www.dhs.gov/isao>. Retrieved on 2015-10-26.
- Fudenberg, Drew, David Levine, Eric Maskin. 1994. The folk theorem with imperfect public information. *Econometrica* **62**(5) 997–1039.
- Fudenberg, Drew, Eric Maskin. 1986. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica: Journal of the Econometric Society* 533–554.
- Gal-Or, Esther, Anindya Ghose. 2005. The economic incentives for sharing security information. *Information Systems Research* **16**(2) 186–208.
- Goehring, Dwight J, James P Kahan. 1976. The uniform n-person prisoner's dilemma game construction and test of an index of cooperation. *Journal of Conflict Resolution* **20**(1) 111–128.

- Gordon, Lawrence A, Martin P Loeb, William Lucyshyn. 2003. Sharing information on computer systems security: An economic analysis. *Journal of Accounting and Public Policy* **22**(6) 461–485.
- Gordon, Lawrence A, Martin P Loeb, William Lucyshyn, Tashfeen Sohail. 2006. The impact of the sarbanes-oxley act on the corporate disclosures of information security activities. *Journal of Accounting and Public Policy* **25**(5) 503–530.
- Gordon, Lawrence A, Martin P Loeb, William Lucyshyn, Lei Zhou. 2015. The impact of information sharing on cybersecurity underinvestment: a real options perspective. *Journal of Accounting and Public Policy* **34**(5) 509–519.
- Kandori, Michihiro. 2002. Introduction to repeated games with private monitoring. *Journal of Economic Theory* **102**(1) 1–15.
- Kandori, Michihiro, Hitoshi Matsushima. 1998. Private observation, communication and collusion. *Econometrica* 627–652.
- Laube, Stefan, Rainer Böhme. 2015. The economics of mandatory security breach reporting to authorities. *Workshop on the economics of information security (WEIS)*.
- Mailath, George J, Stephen Morris. 2002. Repeated games with almost-public monitoring. *Journal of Economic Theory* **102**(1) 189–228.
- Mailath, George J, Larry Samuelson. 2006. *Repeated games and reputations*, vol. 2. Oxford university press Oxford.
- Naghizadeh, Parinaz, Mingyan Liu. 2016a. Inter-temporal incentives in security information sharing agreements. *Position paper for the AAAI Workshop on Artificial Intelligence for Cyber-Security*.
- Naghizadeh, Parinaz, Mingyan Liu. 2016b. Inter-temporal incentives in security information sharing agreements. *Information Theory and Applications Workshop (ITA), 2016*. IEEE.
- Obama, Barack. 2013. Executive order 13636: Improving critical infrastructure cybersecurity. URL <https://www.whitehouse.gov/the-press-office/2013/02/12/executive-order-improving-critical-infrastructure-cybersecurity>. Retrieved on 2015-10-26.
- Obama, Barack. 2015. Executive order 13691: Promoting private sector cybersecurity information sharing. URL <https://www.whitehouse.gov/the-press-office/2015/02/13/executive-order-promoting-private-sector-cybersecurity>. Retrieved on 2015-10-26.
- Ogut, Hulisi, Nirup Menon, Srinivasan Raghunathan. 2005. Cyber insurance and its security investment: Impact of interdependence risk. *WEIS*.
- Park, Jee-Hyeong. 2011. Enforcing international trade agreements with imperfect private monitoring. *The Review of Economic Studies* **78**(3) 1102–1134.
- Romanosky, Sasha, Rahul Telang, Alessandro Acquisti. 2011. Do data breach disclosure laws reduce identity theft? *Journal of Policy Analysis and Management* **30**(2) 256–286.

Sugaya, Takuo. 2011. Folk theorem in repeated games with private monitoring. *Economic Theory Center Working Paper* (011-2011).

Sugaya, Takuo. 2013. Folk theorem in repeated games with private monitoring .

The White House. 2015a. Fact sheet: Administration cybersecurity efforts 2015. URL <https://www.whitehouse.gov/the-press-office/2015/07/09/fact-sheet-administration-cybersecurity-eff>
Retrieved on 2015-10-26.

The White House. 2015b. Summary description of the CNCI. URL <https://www.whitehouse.gov/issues/foreign-policy/cybersecurity/national-initiative>.
Retrieved on 2015-10-26.

Threat Track. 2013. Majority of malware analysts aware of data breaches not disclosed by their employers.
URL <http://www.threattracksecurity.com/press-release/majority-of-malware-analysts-aware-of-data-b>
Retrieved on 2016-02-24.

Verizon. 2015. Vocabulary for event recording and incident sharing. URL <http://veriscommunity.net/index.html>.